Encountered problems when using published fish contamination data in a probabilistic intake assessment

Isabelle Sioen[1], Maaike Bilau, John Van Camp, Stefaan De Henauw

[1]Ghent University, Belgium

Introduction

Marine foods represent a unique source of long chain poly-unsaturated fatty acids (LC-PUFA) of the omega-3 (n-3) family. Moreover, they also contain a number of other valuable nutrients (e.g. high quality amino acids and micronutrients like vitamin D, iron, iodine, selenium and zinc). Today it is generally accepted that fish is important in a healthy and balanced omnivorous human diet. This favourable health perception of marine foods is however troubled by less favourable information regarding the potential adverse health impact of chemical contamination of marine foods. Persistent organochlorine compounds, e.g., PCBs, dioxin-like substances, and organochlorine pesticides (DDT/DDE) accumulate in the marine food chain. On the other hand cadmium, lead, mercury and arsenic are well known marine-environment pollutants. This overall picture forms a potential basis for an important conflictuous model between dietary recommendations and toxicological safety assurance. Most scientists seem to agree that the intake of n-3 LC-PUFA in the average Western diet is currently too low and should substantially increase and that this is difficult to achieve on the basis of the current Western food supply and consumption patterns. A substantial increase in fish consumption could be part of the solution. This would however also increase the total intake of persistent organic chemicals and heavy metals.

Therefore it is useful to carry out a risk-benefit analysis to quantify the conflict linked with fish consumption. One important step in that analysis is a detailed intake assessment of nutrients and contaminants of interest by fish consumption, in order to know precisely the current state-of-art. Special attention will go to the origin of the fish consumed by the Belgian population, since the contamination concentration of fish can depend largely on their origin. The intake assessment will be done with the following simulation model, combining fish consumption data with nutrient and contaminant concentrations:

$$Y_i = \Sigma_v \Sigma_a \Sigma_t (p_f X_{v,i,t} b_{v,a} C_{v,a})/(BW_i T)$$

where Y_i= average daily intake of subject i; $X_{v,i,t}$= amount (g) of fish v consumed by subject i (with body weight BW_i), at day t ($t = 1,\ldots,T$); p_f= effect of processing using preparation method f on $C_{v,a}$; $b_{v,a}$= probability determining whether fish v comes from region a and $C_{v,a}$= concentration of a specific nutrient/contaminant in fish v from region a.

A probabilistic approach will be applied for the simulations. Such an approach is essential to represent the complexity of real situations and takes into account variability and uncertainty [1]. The variability of the different input parameters is taken into account by representing them not as single data points but as probability functions (parametric distributions). This seems very relevant in the case of contaminant concentrations in fish since it is more realistic to represent for example the dioxin concentration of herring in the Baltic Sea per gram herring by a distribution expressing the natural variability than representing it by one single data point (e.g. the mean). The same argument goes for the consumption data, since the high variability in human diet. A key component of a probabilistic intake assessment is however the selection of the most appropriate input distribution to represent exposure variables, such that it represents the variability in a right way [2]. It is recognised that the development of the distributions will rely on the collection of accurate quantitative data [1]. This abstract describes the methods used and the problems encountered when trying to characterize appropriate input distributions to represent the dioxin and PCB concentrations for different fish species of different fishing areas relevant for Belgian consumption.

Materials and Methods

An Excel®-database has been constituted with regard to the concentration in different fish species of different fishing areas for several contaminants, e.g. dioxins and (dioxin-like) PCBs. In the database, all relevant information is included: commercial name, scientific name, farmed or wild fish, period of capture, age of the fish at the moment of analysis, fishing area where the fish was caught, number of samples, number of individuals per sample in the case of pooled samples, fat content of the fish and (mean) contaminant content, with extra statistical data if available (standard deviation, minimum and maximum). At this moment, the database contains data of 3 free available databases (NIFES; ICES; CFSAN), 24 reports of governments/research institutes and 77 peer reviewed research papers. All of them are found using the following resources: PubMed, Web of Science and Google.

Once the data were collected, BestFit® distribution fitting program V.4.5 (Palisade Corp., Newfield, NY, USA) has been used to look for the distribution shape and parameters that describe the data in the best way. This program identifies the distributions that best describe a given dataset using goodness-of-fit tests.

Results and Discussion

When screening publications to build up the database, three important problems were encountered. First, different analytical methodologies can be used for determining the same substance. In the case of (dioxin-like) PCBs, it is not always clear which congeners were analysed. Nevertheless, this is important to know when comparing the overall content reported in different publications. For example, when one wants to compare the data of the salmon study of Hites et al.[3] to other data reporting the PCBcontent in wild and farmed salmon, a good comparison is biased because Hites et al.[3] measured Aroclor 1254 whereas Easton et al.[4] measured 112 undefined PCB congeners. As a result, only the concentration data of publications clearly mentioning that they measured the seven marker PCBs (congeners 28, 52, 101, 118, 138, 153, 180) and/or the twelve non-ortho and mono-ortho dioxin-like PCBs (congeners 77, 81, 126, 169, 105, 114, 118, 123, 156, 157, 167, 189) will be used for the intake assessment. Another encountered problem is that, over the different analyses, some analysed individual samples whereas other used pooled samples. To solve that problem, weighing factors are taken into account, considering the concentration measured in a pooled sample of n individuals as a mean of n individual samples. Other biases between the different results reported are created by different ways of handling the sample prior to the analysis, for example removing the skin (which can create a decrease in the contaminant concentration). And in some cases it is not clear whether or not only edible parts of the fish are analysed. This results in uncertainty about the input data. Moreover, using different methodologies for calculating TEQ-values for dioxins (I-TEQ versus WHO-TEQ), makes data less comparable. Lindstrom et al.[5] already concluded that the world-wide data on dioxins, at background levels in foods, are not direct numerically comparable due to inherent analytical reasons. When comparing interlaboratory results on the determination of dioxins and dioxin-like PCBs in three natural food items, up to 50% deviations were found. For the salmon, sample levels of 5.9 to 38 ppt TEQ (by 25 laboratories) were reported where 15 ppt was the consensus value[5]. So all these interresearch differences create biases when using the different data for fitting one distribution expressing the substance concentration in a certain species. The probabilistic approach makes it possible to take these uncertainties into account by using confidence intervals. But nevertheless, the more unsure the input data, the more unsure the output.

Second, the analytical results reported are expressed in many different ways. Some authors report all data of individual measurements, whereas others only report mean values of a certain amount of samples (sometimes with standard deviation and/or range). To solve this problem, weighing factors are calculated according to the number of samples on which the reported value is based. According to the weighing factors and the reported concentrations, cumulative probability functions are created using BestFit®, expressing the probability for all possible concentration of a certain contaminant in a certain fish species. On the bases of the graphs and the test statistics created by BestFit®, it has to be decided which kind of distribution and which parameters can be used to describe the data as good as possible and to use finally as model input. In this approach, having a lot of individual measurements would be the best in order to create a good distribution expressing the variability in the concentrations. Another problem is that concentrations can be expressed per gram fish or per gram fat. When the fat content is not reported in the latest case, there is no possibility to make those data comparable and to use them in this approach.

Third, there is a lack of data. For different species no contamination data can be found for commercially important regions; e.g. no dioxin concentration data are yet found for halibut of the Northwest Atlantic Sea, although more than 55% of the halibut on the Belgian market originates from that fishing area. Therefore, some species have to be grouped (in the case of dioxins and PCBs according to fat content) per fishing area, or some sub areas have to be

grouped to bigger areas. Since a lot of governmental publications report data without mentioning the origin of the fish (because in most of the cases samples of the local market were analysed), they are difficult to use in our approach, since we aimed to distinguish contamination data from different regions. This problem is partly solved, using the FAO FISHSTAT Plus data[6] to discover the origin and assuming that the fish bought on the local market is caught by the country itself. In addition, in several ecological or monitoring studies, it is not clear whether the samples analysed are representative for human consumption. Therefore, more detailed information about the fish samples would be very useful in order to use them in intake and risk assessment approaches.

It can be concluded that several problems are encountered when building up databases with contaminant concentrations in fish, related to a lack of international regulations about the analytical methodologies to be used and the manner of reporting the results. This study confirms a statement earlier made by the EU Scientific Steering Committee about the need to improve the comparability of data critical for the conduct of exposure assessment [1].

Acknowledgements

The authors gratefully acknowledge financial support from the Belgian Science Policy through the SPSD II project CP/02/56 and the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). They also want to thank Frederik Verdonck for statistical advice.

References

1. EU Scientific Steering Committee (2000), First report on the harmonisation of risk assessment procedures in the European Union.

2. Gilsenan M. B., Lambe J. and Gibney M. J. (2003) *Food Addit Contam* 20: 1023-1033.

3. Hites R. A., Foran J. A., Carpenter D. O., Hamilton M. C., Knuth B. A. and Schwager S. J. (2004) *Science* 303: 226-229.

4. Easton M. D. L., Luszniak D. and der Geest E. (2002) *Chemosphere* 46: 1053-1074.

5. Lindstrom G., Haug L. S., Nicolaysen T. and Dybing E. (2002) *Chemosphere* 47: 139-146.

6. FAO FISHSTAT Plus Version 2.3. Universal software for fishery statistical time series. http://www.fao.org/fi/statist/fisoft/FISHPLUS.asp